KEY

Chapter 8 Review

Problem 1: You are collecting data for the EPA and need to test the concentration of a substance in samples of the ground water throughout different locations. Determine which type of sampling method is used based on each of the following scenarios.

a) You collected data from 50 locations over a week and want to test every 5th sample.

Systematic

b) You need to collect data quickly and only sample locations from the region you are currently in.

Convenience

c) You believe each region will have different concentrations. So you randomly select 6 locations to sample from for each region.

GROUP by Regions + randomly sample within (which are distinct)
⟹ Stratified

d) There are 50 possible locations to sample from and you believe there are no differences between locations. So you randomly select 10 locations to sample from.

Random sample

e) Each region has a diverse locations in terms of concentration. So you randomly select 3 regions and sample from each possible location within the region.

GROUP by regions + census all locations within (which are similar)
⟹ cluster

Problem 2: The following table represents a grouped frequency distribution of the number of hours spent on the computer per week for 50 students.

Hours	Number of Students
0.0–3.4	2
3.5–6.9	19
7.0–10.4	14
10.5–13.9	11
14.0–17.4	4
Total	

a) How many students use the computer less than 7 hours per week?

$2 + 19 = 21$ students

b) What percent of students used the computer more than 10.4 hours per week?

$\frac{11+4}{50} = \frac{15}{50} =$ 30 %

c) What percent of students used the computer between 7.0 and 13.9 hours per week inclusive?

$\frac{14+11}{50} = \frac{25}{50} =$ 50 %

Problem 3: Create a frequency distribution for the following data on students' favorite color:

Yellow, Blue, Red, Red, Blue, Yellow, Green, Green, Blue, Red, Yellow, Yellow, Green, Blue, Green, Blue, Red, Yellow, Yellow, Yellow, Blue, Red, Red, Blue.

Color	Frequency (count)
Yellow	7
Blue	7
Green	4
Red	6

Problem 4: Below is a dataset about the age of a giraffe and its height in meters.

Age	0.5	1.5	1	2	4	6	8	12	2	4
Height (m)	5	8	7	9	8	8	9	10	6	9

a) Calculate the correlation for the dataset above and determine if it is statistically significant at a level of significance of $\alpha = 0.05$.

→ LinReg (ax + b)

X = Age, Y = Height

OR → 2-Var Stats (X = Age, Y = Height)

→ $r = 0.7176$ $n = 10$ $\alpha = 0.05$

→ $|r| = 0.7176 > 0.632 = CV$

⇒ Significant

b) If appropriate, determine the regression equation.

↳ yes because correlation is significant

$\hat{y} = ax + b = 0.298x + 6.678$

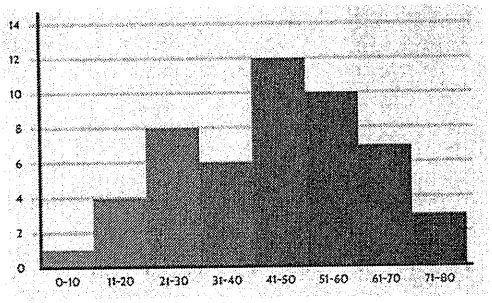
c) If a giraffe is 3.5 years old, make a prediction for how tall it will be.

$\hat{y} = 0.298(3.5) + 6.678 = 7.721$ m

d) If a giraffe is 7 years old, make a prediction for how tall it will be.

$\hat{y} = 0.298(7) + 6.678 = 8.764$ m

Problem 5: The histogram to the right represents ages of attendees at a school fair.



a) What age group had the highest frequency?

41–50 → 12

b) What is the frequency for the 61–70 year-old age group?

7

c) How many attendees were between the ages of 11 and 40 inclusive?

ages: 11–20, 21–30, 31–40

4 + 8 + 6 = 18

d) Which three age groups were the least represented at the fair?

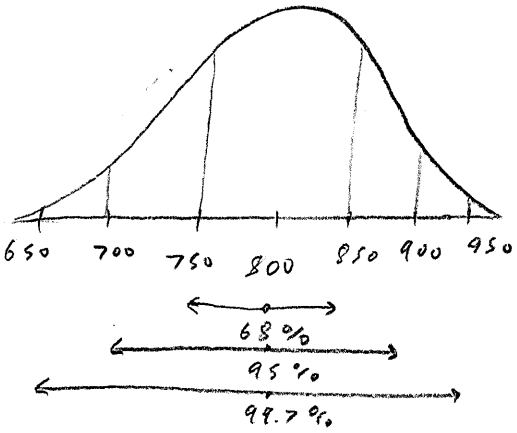
Age	Count
0–10	→ 1
71–80	→ 3
11–20	→ 4

Problem 6: Daily ticket sales for the local zoo have a normal distribution with mean \$800 and standard deviation \$50. Use the empirical rule to answer the following questions.

(☆ Draw, label + shade curve

a) What percent of days have ticket sales between \$750 and \$850?

1 step ⇒ (68 %)



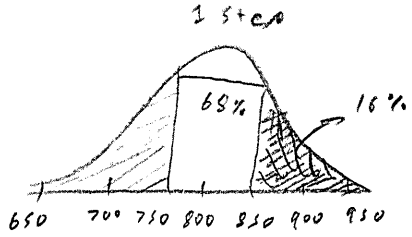
b) 99.7% of days have ticket sales between which two dollar amounts?

↳ 3 steps ⇒ (650, 950)

c) What percent of days have ticket sales between \$700 and \$900?

2 steps ⇒ (95 %)

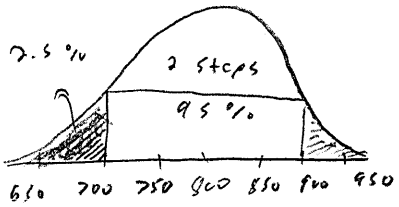
d) What percent of days have ticket sales greater than \$850?



→ outside = Total 100 % − Inside 68 % = 32 %

→ ONLY RIGHT = 32 % / 2 = (16 %)

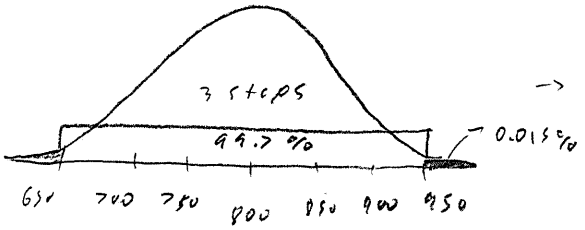
e) What percent have days have ticket sales less than \$700?



→ outside = Total 100 % − Inside 95 % = 5 %

→ ONLY LEFT = 5 % / 2 = (2.5 %)

f) What percent of days have ticket sales greater than \$950?



→ outside = Total 100 % − Inside 99.7 % = 0.3 %

→ ONLY LEFT = 0.03 % / 2 = (0.15 %)

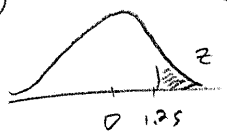
Problem 7: Rainfall per year in a country has a normal distribution with mean 40 inches and standard deviation 4 inches.

a) Find the z-score for a city that has yearly rainfall of 45 inches.

① ★ Find $z = \frac{x-\mu}{\sigma}$ ⟹ ② ★ Draw, label + shade curve ③ ★ z-table

① $z = \frac{45-40}{4} = 1.25$

b) Find the probability a city has yearly rainfall more than 45 inches. ⟹ ★ use z-score just found!

② 

③ ★ Right = 1 − Left

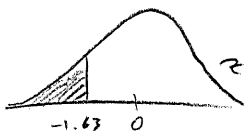
$P(z > 1.25) = 1 - P(z < 1.25)$

$= 1 - 0.8944$

$= 0.1056$

c) Find the probability a city has yearly rainfall less than 33.5 inches.

① $z = \frac{33.5-40}{4} = -1.625$ round 2 decimals ⟶ ≈ -1.63

② 

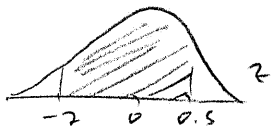
③ ★ Left = Table

$P(z < -1.63) = 0.0516$

d) Find the probability a city has yearly rainfall between 32 and 42 inches per year.

① $z_1 = \frac{32-40}{4} = -2.00$

$z_2 = \frac{42-40}{4} = 0.50$

② 

③ ★ Between = Left z_2 − Left z_1

$= P(z < 0.50) - P(z < -2.00)$

$= 0.6915 - 0.0228$

$= 0.6687$

Problem 8: Here are ages from a sample of attendees at the school fair: 5, 23, 38, 11, 4, 44, 57, 11, 28, 38, 45, 60, 8, 11.

→ n = 14

a) Find the mean age. ★ using calc → 1-Var Stats

$\bar{x} = 27.36$

By hand: $\bar{x} = \frac{5+23+38+\cdots+11}{14} = 27.36$

b) Find the median age.

Med = 25.5

By hand: order → then narrow down to middle

4, 5, 8, 11, 11, 11, 23, 28, 38, 38, 44, 45, 57, 60

Med: $\frac{23+28}{2} = 25.5$

c) Find the mode of ages.

11 → appears 3 times

d) Find the range of ages.

Range = Max − Min = 60 − 4 = 56

e) Find the sample standard deviation of ages.

using calc → $s_x = 19.61$

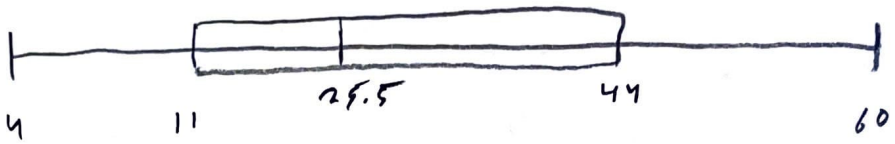
f) Calculate the 5-number summary of ages.

By hand: 4, 5, 8, (11), 11, 11, 23, 28 | 38, 38, (44), 45, 55, 60

min = 4 Q_1 = 11 Med = 25.5 Q_3 = 44 max = 60

By calc: 1-var stats (L_1 = #) → min = 4, Q_1 = 11, Q_3 = 44, Med = 25.5, max = 60

g) Sketch a boxplot of ages based on your answers from (f).



Problem 9: Two new species of walruses were discovered, one that is pink and one that is blue. From a sample of 15 of each kind, the following information was recorded and/or calculated.

- Pink walruses: Sample mean weight $\bar{x} = 1{,}100$ lbs with standard deviation 200 lbs. Suppose the margin of error using a 95% level of confidence is $E = 275$ lbs.
- Blue walruses: Sample mean weight $\bar{x} = 1{,}300$ lbs with standard deviation 150 lbs. Suppose the margin of error using an 95% level of confidence is $E = 215$ lbs.

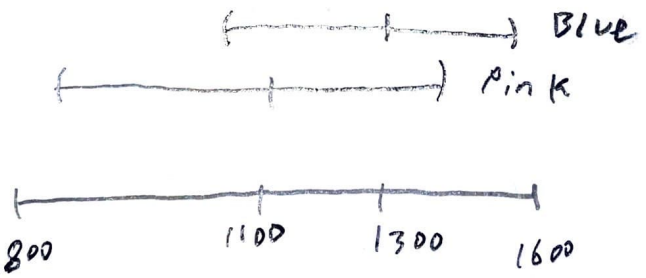
Construct 95% confidence intervals for both types of species (and plot them). Can we conclude that one species is heavier than the other? Why or why not?

Pink: PE ± E = 1100 ± 275 = [825, 1375]

Blue: PE ± E = 1300 ± 215 = [1085, 1515]

→ Intervals overlap

⇒ Can't conclude one is greater



Problem 10: Suppose a sample of 975 spiders were asked if they prefer to dine on flies or butterflies and 600 responded "Butterflies are way better". Calculate the point estimate to estimate the true proportion of spiders who prefer butterflies and construct a 95% confidence using the rule of thumb to calculate the margin of error.

$E = \frac{1}{\sqrt{n}} \times 100\% = \frac{1}{\sqrt{975}} * 100 = 3.20\%$

$\hat{p} = \frac{x}{n} = \frac{600}{975} = 0.6154 \Rightarrow 61.54\%$

$\Rightarrow CI = PE \pm E$

$= 61.54 \pm 3.20$

$= [58.34, 64.74]$